arXiv:0709.0538v1 [cond-mat.stat-mech] 4 Sep 2007

Markov-Chain Monte Carlo Methods for Simulations of Biomolecules

Bernd A. Berg[1,2]

[1] School of Computational Science (SCS), Florida State University, Tallahassee, FL 32306-4130, USA berg@scs.fsu.edu
[2] Department of Physics, Florida State University, Tallahassee, FL 32306-4350, USA berg@hep.fsu.edu

The computer revolution has been driven by a sustained increase of computational speed of approximately one order of magnitude (a factor of ten) every five years since about 1950. In natural sciences this has led to a continuous increase of the importance of computer simulations. Major enabling techniques are Markov Chain Monte Carlo (MCMC) and Molecular Dynamics (MD) simulations.

This article deals with the MCMC approach. First basic simulation techniques, as well as methods for their statistical analysis are reviewed. Afterwards the focus is on generalized ensembles and biased updating, two advanced techniques, which are of relevance for simulations of biomolecules, or are expected to become relevant with that respect. In particular we consider the multicanonical ensemble and the replica exchange method (also known as parallel tempering or method of multiple Markov chains).

1 Introduction

Markov chain Monte Carlo (MCMC) calculations started in earnest with the 1953 paper by Nicholas Metropolis, Arianna Rosenbluth, Marshall Rosenbluth, Augusta Teller and Edward Teller [1]. Since then MCMC simulations have become an indispensable tool with applications in many branches of science. Some of those are reviewed in the proceedings [2] of the 2003 Los Alamos conference, which celebrated the 50th birthday of Metropolis simulations.

The purpose of this article is to give a concise overview ranging from statistical preliminaries and plain Monte Carlo (MC) calculations over basic techniques for MCMC simulations to advanced methods, which are indispensable when it comes to the simulations of complex systems with frustrated interactions. For such systems rugged free energy landscapes are typical. Here our focus is on biomolecules such as peptides (small proteins) in an all-atom approach, defined by a model energy function. At a given temperature this

energy function determines in principle the Gibbs ensemble of the molecule. In practice equilibrium is sometimes hard to reach in MCMC simulations of the canonical ensemble. Considerable improvements can be made by using generalized ensembles and biased sampling.

The first part of this article gives a treatment of the MCMC fundamentals that is largely based on the author's book [3] on the subject, which gives many more details and contains extensive additional material. The book comes with computer code that can be downloaded from the web. The solutions of numerous numerical assignments, are reproducible by compiling and running the corresponding computer programs. Informations and a link to the computer code are found on the web at `http://www.scs.fsu.edu/~berg`.

The second part of this article builds on original literature of generalized ensemble methods [4, 5, 6, 7, 8, 9, 10, 11, 12]. We start with a brief history and elaborate on the underlying ideas. Subsequently we turn to biophysics, where generalized ensembles were introduced in Ref. [13, 14]. Finally a scheme for biasing the Metropolis updating proposals [15] is considered, which can be combined with generalized ensembles.

As our emphasize is on explaining MCMC methods, we restrict ourselves to a simply models, which are well suited for illustrating the essence of a method. Our article is organized as follows. The next section introduces the MCMC method. Section 3 discusses the statistical analysis of autocorrelated MCMC data. Section 4 deals with generalized ensembles and section 5 with biased MCMC updating. A short outlook and conclusions are given in the final section 6. It should be noted that these lecture notes are not supposed to be an unbiased review, but are based on work in which the author has been involved our is particularly interested.

2 Markov Chain Monte Carlo

2.1 Statistical preliminaries

Let $f(x)$ be a probability density and x^r its associated random variable. The **(cumulative) distribution function** of the random variable x^r is defined as

$$F(x) = P(x^r \le x) = \int_{-\infty}^{x} f(x)\, dx \tag{1}$$

where $P(x^r \le x)$ is the probability for $x^r \le x$. A particularly simple and important case is the **uniform probability distribution** for random numbers between $[0, 1)$,

$$u(x) = \begin{cases} 1 & \text{for } \ 0 \le x < 1; \\ 0 & \text{elsewhere.} \end{cases} \tag{2}$$

Remarkably, the uniform distribution allows for the construction of general probability distributions. Let

$$y = F(x) = \int_{-\infty}^{x} f(x')\, dx'$$

and assume that the inverse $x = F^{-1}(y)$ exists. For y^r being a uniformly distributed random variable in the range $[0, 1)$ it follows that

$$x^r = F^{-1}(y^r) \tag{3}$$

is distributed according to the probability density $f(x)$. To generate the uniform distribution on a computer, one relies on pseudo random number generators. Desirable properties are randomness according to statistical tests, a long period, computational efficiency, repeatability, portability, and homogeneity (all subsets of bits are random). Our purposes are served well by the generator of Marsaglia and collaborators [21], which comes as part of the code of [3].

2.2 Partition function and Potts models

MC simulations of systems described by the Gibbs canonical ensemble aim at calculating estimators of physical observables at a temperature T. In the following we choose units so that the Boltzmann constant becomes one, i.e. $\beta = 1/T$. Let us consider the calculation of the **expectation value** of an **observable** $\mathcal{O}$. Mathematically all systems on a computer are discrete, because a finite word length has to be used. Hence, the expectation value is given by the sum

$$\widehat{\mathcal{O}} = \widehat{\mathcal{O}}(\beta) = \langle \mathcal{O} \rangle = Z^{-1} \sum_{k=1}^{K} \mathcal{O}^{(k)}\, e^{-\beta\, E^{(k)}} \tag{4}$$

$$\text{where} \quad Z \;=\; Z(\beta) = \sum_{k=1}^{K} e^{-\beta\, E^{(k)}} \tag{5}$$

is the **partition function**. The index $k = 1, \ldots, K$ labels the **configurations** of the system, and $E^{(k)}$ is the (internal) energy of configuration k. The configurations are also called **microstates**. To distinguish the configuration index from other indices, it is put in parenthesis.

We introduce generalized Potts models on d-dimensional hypercubic lattices with periodic boundary conditions (i.e., the models are defined on a torus in d dimensions). Without being overly complicated, these models are general enough to illustrate the essential features of MCMC simulations. Various subcases are by themselves of physical interest. We define the energy function of the system by

$$E^{(k)} = -2 \sum_{\langle ij \rangle} \delta(q_i^{(k)}, q_j^{(k)}) + \frac{2\, d\, N}{q} \quad \text{where} \quad \delta(q_i, q_j) = \begin{cases} 1 \text{ for } q_i = q_j \\ 0 \text{ for } q_i \neq q_j\,. \end{cases} \tag{6}$$

The sum $\langle ij \rangle$ is over the nearest neighbor lattice sites and $q_i^{(k)}$ is called the **Potts spin** or **Potts state** of configuration k at site i. For the q-state Potts

model, $q_i^{(k)}$ takes on the values $1, \ldots, q$. The case $q = 2$ becomes equivalent to the Ising ferromagnet. See F.Y. Wu [16] for a review of Potts models. In $d = 2$ dimensions the phase transition is second order for $q \leq 4$ and first order for $q \geq 5$. The exact infinite volume latent heats $\triangle e_s$ and entropy discontinuities $\triangle s$ were calculated by Baxter [17], while the interface tensions f_s were derived later, see [18] and references therein.

2.3 Sampling, reweighting, and important configurations

For the Ising model (2-state Potts) it is straightforward to **sample statistically independent configurations**. We simply have to generate N spins, each either 0 or 1 with 50% likelihood. This is called **random sampling**. In Fig. 1 a thus obtained histogram for the $2d$ Ising model **energy per spin** is depicted.

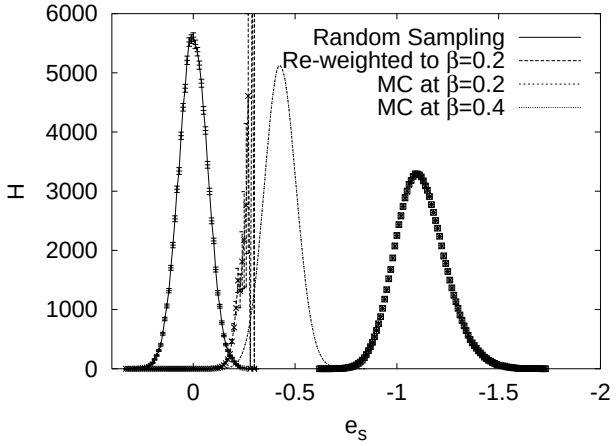

Fig. 1. Energy histograms of 100 000 entries each for the Ising model on a 20×20 lattice: Random Sampling gives statistically independent configurations at $\beta = 0$. Histograms at $\beta = 0.2$ and $\beta = 0.4$ are generated with Markov chain MC. Reweighting of the $\beta = 0$ random configurations to $\beta = 0.2$ is shown to fail. These are assignments `a0301_02` and `a0303_02` of [3].

Note that it is important to distinguish the energy measurements on single configurations from the expectation value. The expectation value $\widehat{e}_s$ is a single number, while e_s fluctuates. From the measurement of many e_s values one finds an estimator of the mean, $\overline{e}_s$, which fluctuates with a reduced variance.

The histogram entries at $\beta = 0$ can be **reweighted** so that they correspond to other β values. We simply have to multiply the entry corresponding to energy E by $\exp(-\beta E)$. Similarly histograms corresponding to the Gibbs ensemble at some value β_0 can be reweighted to other β values. Care has to

be taken to ensure that the involved arguments of the exponential function do not become too large [3]. Reweighting has a long history, which we discuss in section 4.1.

In Fig. 1 reweighting is done from $\beta_0 = 0$ to $\beta = 0.2$. But, by comparison to the histogram from a Metropolis MC calculation at $\beta = 0.2$, the result is seen to be disastrous. The reason is easily identified: In the range where the $\beta = 0.2$ histogram takes on its maximum, the $\beta = 0$ histogram has not a single entry. Our random sampling procedure misses the important configurations at $\beta = 0.2$. Reweighting to new β values works only in a range $\beta_0 \pm \triangle\beta$, where $\triangle\beta \to 0$ in the infinite volume limit. Details are given in section 4.1.

Let us determine the important contributions to the partition function. The partition function can be re-written as a sum over energies

$$Z = Z(\beta) = \sum_E n(E)\, e^{-\beta\, E} \tag{7}$$

where the unnormalized spectral density $n(E)$ is defined as the number of microstates k with energy E. For a fixed value of β the energy probability density

$$P_\beta(E) = c_\beta\, n(E)\, e^{-\beta E} \tag{8}$$

is peaked around the average value $\widehat{E}(\beta)$, where c_β is a normalization constant determined by $\sum_E P_\beta(E) = 1$. The important configurations at temperature $T = 1/\beta$ are at the energy values for which the probability density $P_\beta(E)$ is large. To sample them efficiently, one needs a procedure which generates the configurations with their Boltzmann weights

$$w_B^{(k)} = e^{-\beta E^{(k)}}\ . \tag{9}$$

The number of configurations $n(E)$ and the weights combine then so that the probability to generate a configuration at energy E becomes $P_\beta(E)$ as given by equation (8).

2.4 Importance sampling and Markov chain Monte Carlo

For the canonical ensemble **importance sampling** generates configurations k with probability

$$P_B^{(k)} = c_B\, w_B^{(k)} = c_B\, e^{-\beta E^{(k)}} \tag{10}$$

where the constant c_B is determined by the normalization condition $\sum_k P_B^{(k)} = 1$. The vector $(P_B^{(k)})$ is called **Boltzmann state**. When configurations are stochastically generated with probability $P_B^{(k)}$, the expectation value becomes the **arithmetic average:**

$$\widehat{\mathcal{O}} = \widehat{\mathcal{O}}(\beta) = \langle \mathcal{O} \rangle = \lim_{N_K \to \infty} \frac{1}{N_K} \sum_{n=1}^{N_K} \mathcal{O}^{(k_n)}\ . \tag{11}$$

Truncating the sum at some finite value of N_K, we obtain an **estimator of the expectation value**

$$\overline{\mathcal{O}} = \frac{1}{N_K} \sum_{n=1}^{N_K} \mathcal{O}^{(k_n)} \ . \tag{12}$$

Normally, we cannot generate configurations k directly with the probability (10), but they may be found as members of the equilibrium distribution of a dynamic process. A **Markov process** is a particularly simple dynamic process, which generates configuration k_{n+1} stochastically from configuration k_n, so that no information about previous configurations $k_{n-1}, k_{n-2}, \ldots$ is needed. The elements of the Markov process **time series** are the configurations. Assume that the configuration k is given. Let the transition probability to create the configuration l in one step from k be given by $W^{(l)(k)} = W[k \to l]$. The **transition matrix**

$$W = \left(W^{(l)(k)}\right) \tag{13}$$

defines the Markov process. Note, that this matrix is a very big (never stored in the computer), because its labels are the configurations. To generate configurations with the desired probabilities, the matrix W needs to satisfy the following properties:

(i) **Ergodicity**:

$$e^{-\beta E^{(k)}} > 0 \ \text{ and } \ e^{-\beta E^{(l)}} > 0 \quad \text{imply :} \tag{14}$$

an integer number $n > 0$ exists so that $(W^n)^{(l)(k)} > 0$ holds.

(ii) **Normalization**:

$$\sum_l W^{(l)(k)} = 1 \ . \tag{15}$$

(iii) **Balance**:

$$\sum_k W^{(l)(k)}\, e^{-\beta E^{(k)}} \ = \ e^{-\beta E^{(l)}} \ . \tag{16}$$

The Boltzmann state (10) is an eigenvector with eigenvalue 1 of the matrix $W = (W^{(l)(k)})$.

An **ensemble** is a collection of configurations for which to each configuration k a probability $P^{(k)}$ is assigned, $\sum_k P^{(k)} = 1$. The **Gibbs or Boltzmann ensemble** E_B is defined to be the ensemble with the probability distribution (10).

An **equilibrium ensemble** E_{eq} of the Markov process is defined by its probability distribution P_{eq} satisfying

$$W\, P_{eq} = P_{eq}\,, \ \text{ in components } \ P_{eq}^{(l)} = \sum_k W^{(l)(k)} P_{eq}^{(k)} \ . \tag{17}$$

Statement: Under the conditions (i), (ii) and (iii) the Boltzmann ensemble is the **only** equilibrium ensemble of the Markov process and an **attractive fixed point**. Applying the transition matrix n times give rise to an ensemble E^n. For $n \to \infty$ the distance between E^n and the Boltzmann ensemble decreases asymptotically like

$$||E^n - E_B|| \leq \exp(-\lambda\, n)\, ||E^0 - E_B|| \tag{18}$$

where E^0 is the initial ensemble and $\lambda > 0$ a constant.

For a proof the readers are referred to [3]. There are many ways to construct a Markov process satisfying (i), (ii) and (iii). A stronger condition than balance (16) is

(iii') **Detailed balance**:

$$W^{(l)(k)}\, e^{-\beta E^{(k)}} \;=\; W^{(k)(l)} e^{-\beta E^{(l)}} \;. \tag{19}$$

Using the normalization $\sum_k W^{(k)(l)} = 1$ detailed balance implies balance.

At this point we have succeeded to replace the canonical ensemble average by a time average over an artificial dynamics. Calculating averages over large times, like one does in real experiments, is equivalent to calculating averages of the ensemble.

2.5 Metropolis and heatbath algorithm for Potts models

The **Metropolis algorithm** can be used whenever one knows how to calculate the energy of a configuration. Given a configuration k, the Metropolis algorithm proposes a configuration l with probability

$$f(l,k) \;\; \text{normalized to} \quad \sum_l f(l,k) = 1 \;. \tag{20}$$

The new configuration l is accepted with probability

$$w^{(l)(k)} = \min\left[1,\, \frac{P_B^{(l)}}{P_B^{(k)}}\right] = \begin{cases} 1 \;\; \text{for} \;\; E^{(l)} < E^{(k)} \\ \exp[-\beta(E^{(l)} - E^{(k)})] \;\; \text{for} \;\; E^{(l)} > E^{(k)}. \end{cases} \tag{21}$$

If the new configuration is rejected, the old configuration has to be counted again. The **acceptance rate** is defined as the ratio of accepted changes over proposed moves. With this convention we do not count a move as accepted when it proposes the at hand configuration.

The Metropolis procedure gives rise to the transition probabilities

$$W^{(l)(k)} = f(l,k)\, w^{(l)(k)} \;\; \text{for} \;\; l \neq k \tag{22}$$

$$\text{and} \;\; W^{(k)(k)} = f(k,k) + \sum_{l \neq k} f(l,k)\, (1 - w^{(l)(k)}) \;. \tag{23}$$

Therefore, the ratio $\left(W^{(l)(k)}/W^{(k)(l)}\right)$ satisfies detailed balance (19) if

$$f(l,k) \;=\; f(k,l) \;\text{ holds}\,. \tag{24}$$

Otherwise the probability density $f(l,k)$ is unconstrained. So there is an amazing flexibility in the choice of the transition probabilities $W^{(l)(k)}$. Also, the algorithm generalizes immediately to arbitrary weights.

If sites are chosen with the uniform probability distribution $1/N$ per site, where N is the total number of spins, it is obvious that the algorithm fulfills detailed balance. It is noteworthy that the procedure remains valid when the spins are chosen in the systematic order $1,\ldots,N$. Balance (16) still holds, whereas detailed balance (19) is violated (an exercise of Ref. [3]).

If one performs multiple hits with the Metropolis algorithm at the same spin (**multi-hit Metropolis algorithm**), the local Boltzmann distribution defined by its nearest neighbors is approached for an increasing number of hits. The **heatbath algorithm (HBA)** chooses a state q_i directly with the local Boltzmann distribution defined by its nearest neighbors. The state q_i can take on one of the values $1,\ldots,q$ and, with all other states set, determines a value of the energy function (6). We denote this energy by $E(q_i)$ and the Boltzmann probabilities are

$$P_B(q_i) \;=\; \text{const}\; e^{-\beta\, E(q_i)} \tag{25}$$

where the constant is determined by the normalization condition

$$\sum_{q_i=1}^{q} P_B(q_i) \;=\; 1\;. \tag{26}$$

In equation (25) we can define $E(q_i)$ to be just the contribution of the interaction of q_i with its nearest neighbors to the total energy and absorb the other contributions into the overall constant. Here we give a generic HBA which works for arbitrary values of q and d (other solutions can be more efficient). We calculate the cumulative distribution function of the heatbath probabilities

$$P_{HB}(q_i) = \sum_{q_i'=1}^{q_i} P_B(q_i')\;. \tag{27}$$

The normalization condition (26) implies $P_{HB}(q) = 1$. Comparison of these cumulative probabilities with a uniform random number x^r yields the heatbath update $q_i \to q_i'$. Note that in the heatbath procedure the original value $q_i^{\rm in}$ does not influence the selection of $q_i^{\rm new}$.

2.6 Heatbath algorithm for a continuous system

We give the $O(3)$ σ model as an example of a model with a continuous energy function. Expectation values are calculated with respect to the partition function

$$Z \;=\; \int \prod_i ds_i \; e^{-\beta E(\{s_i\})} \quad \text{of spins} \;\; \boldsymbol{s}_i = \begin{pmatrix} s_{i,1} \\ s_{i,2} \\ s_{i,3} \end{pmatrix} \tag{28}$$

which are normalized to lie on the unit sphere, $(\boldsymbol{s}_i)^2 = 1$. The measure ds_i is defined by

$$\int ds_i \;=\; \frac{1}{4\pi} \int_{-1}^{+1} d\cos(\theta_i) \int_0^{2\pi} d\phi_i \;, \tag{29}$$

where the polar (θ_i) and azimuth (ϕ_i) angles define the spin s_i on the unit sphere. The energy is

$$E \;=\; -\sum_{\langle ij \rangle} \boldsymbol{s}_i \boldsymbol{s}_j \;, \tag{30}$$

where the sum goes over the nearest neighbor sites of the lattice and $\boldsymbol{s}_i \boldsymbol{s}_j$ is the dot product of the vectors. The $2d$ version of the model is of interest to field theorists because of its analogies with the four-dimensional Yang-Mills theory. In statistical physics the d-dimensional model is known as the **Heisenberg ferromagnet** (references can be found in [3]).

We would like to update a single spin $\boldsymbol{s}$. The sum of its $2d$ neighbors is

$$\boldsymbol{S} = \boldsymbol{s}_1 + \boldsymbol{s}_2 + \ldots + \boldsymbol{s}_{2d-1} + \boldsymbol{s}_{2d} \;.$$

Hence, the contribution of spin $\boldsymbol{s}$ to the energy is $2d - \boldsymbol{s}\boldsymbol{S}$. We may propose a new spin $\boldsymbol{s}'$ with the measure (29) by drawing two uniformly distributed random numbers

$$\begin{aligned} \phi^r \in [0, 2\pi) \quad & \text{for the azimuth angle and} \\ \cos(\theta^r) = x^r \in [-1, +1) \quad & \text{for the cosine of the polar angle}\,. \end{aligned}$$

This defines the probability function $f(\boldsymbol{s}', \boldsymbol{s})$ of the Metropolis process, which accepts the proposed spin $\boldsymbol{s}'$ with probability

$$w(\boldsymbol{s} \to \boldsymbol{s}') \;=\; \begin{cases} 1 \;\; \text{for} \;\; \boldsymbol{S}\boldsymbol{s}' > \boldsymbol{S}\boldsymbol{s}\,, \\ \exp[-\beta(\boldsymbol{S}\boldsymbol{s} - \boldsymbol{S}\boldsymbol{s}')] \;\; \text{for} \;\; \boldsymbol{S}\boldsymbol{s}' < \boldsymbol{S}\boldsymbol{s}\,. \end{cases}$$

One would prefer to choose $\boldsymbol{s}'$ directly with the probability

$$W(\boldsymbol{s} \to \boldsymbol{s}') \;=\; P(\boldsymbol{s}'; \boldsymbol{S}) \;=\; \text{const}\, e^{\beta\, \boldsymbol{s}'\boldsymbol{S}} \;.$$

The HBA creates this distribution. Implementation of it becomes feasible when the energy function allows for an explicit calculation of the probability $P(\boldsymbol{s}'; \boldsymbol{S})$. This is an easy task for the $O(3)$ σ-model. Let

$$\alpha = \text{angle}(\boldsymbol{s}', \boldsymbol{S}), \quad x = \cos(\alpha) \;\; \text{and} \;\; S = \beta|\boldsymbol{S}| \;.$$

For $S = 0$ a new spin $\boldsymbol{s}'$ is simply obtained by random sampling. We assume in the following $S > 0$. The Boltzmann weight becomes $\exp(xS)$ and the normalization constant follows from

$$\int_{-1}^{+1} dx\, e^{xS} \;=\; \frac{2}{S}\,\sinh(S)\;.$$

Therefore, the desired probability is

$$P(\boldsymbol{s}^{'};\boldsymbol{S}) \;=\; \frac{S}{2\sinh(S)}\,e^{xS} \;=:\, f(x)$$

and Eq. (3) can be used to generate events with the probability density $f(x)$. A uniformly distributed random number $y^r \in [0,1)$ translates into

$$x^r = \cos\alpha^r = \frac{1}{S}\ln\left[\,\exp(+S) - y^r\,\exp(+S) + y^r\,\exp(-S)\right]\;. \qquad (31)$$

To give $\boldsymbol{s}^{'}$ a direction in the plane orthogonal to $\boldsymbol{S}$, one chooses a uniformly distributed angle β^r in the range $0 \le \beta^r < 2\pi$. Then, $x^r = \cos\alpha^r$ and β^r completely determine $\boldsymbol{s}^{'}$ with respect to $\boldsymbol{S}$. Before storing $\boldsymbol{s}^{'}$ in the computer memory, we have to calculate coordinates of $\boldsymbol{s}^{'}$ with respect to a Cartesian coordinate system, which is globally used for all spins of the lattice. This amounts to a linear transformation.

3 Statistical Errors of MCMC Data

In large scale MC simulation it may take months, possibly years, to collect the necessary statistics. For such data a thorough error analysis is a must. A typical MC simulation falls into two parts:

1. **Equilibration:** Initial sweeps are performed to reach the equilibrium distribution. During these sweeps measurements are either not taken at all or they have to be discarded when calculating equilibrium expectation values.
2. **Data Production:** Sweeps with measurements are performed. Equilibrium expectation values are calculated from this statistics.

A rule of thumb is: **Do not spend more than 50% of your CPU time on measurements.** The reason for this rule is that one cannot be off by a factor worse than two ($\sqrt{2}$ in the statistical error).

How many sweeps should be discarded for reaching equilibrium? In some situations this question can be rigorously answered with the *Coupling from the Past* method [19] (for a review see [20]). The next best thing to do is to measure the integrated autocorrelation time and to discard, after reaching a visually satisfactory situation, a number of sweeps which is larger than the integrated autocorrelation time. In practice even this can often not be achieved.

Therefore, it is re-assuring that it is sufficient to pick the number of discarded sweeps approximately right. With increasing statistics the contribution

of the non-equilibrium data dies out like $1/N$, where N is the number of measurements. This is eventually swallowed by the statistical error, which declines only like $1/\sqrt{N}$. The point of discarding the equilibrium configurations is that the factor in front of $1/N$ can be large.

There can be far more involved situations, like that the Markov chain ends up in a metastable configuration, which may even stay unnoticed (this tends to happen in complex systems like spin glasses or proteins).

3.1 Autocorrelations

We like to estimate the expectation value $\widehat{f}$ of some physical observable. We assume that the system has reached equilibrium. How many MC sweeps are needed to estimate $\widehat{f}$ with some desired accuracy? To answer this question, one has to understand the autocorrelations within the Markov chain.

Given is a **time series** of N measurements from a Markov process

$$f_i = f(x_i), \quad i = 1, \ldots, N \ , \tag{32}$$

where x_i are the configurations generated. The label $i = 1, \ldots, N$ runs in the temporal order of the Markov chain and the elapsed time (measured in updates or sweeps) between subsequent measurements f_i, f_{i+1} is always the same. The estimator of the expectation value $\widehat{f}$ is

$$\overline{f} = \frac{1}{N} \sum f_i \ . \tag{33}$$

With the notation

$$t = |i - j|$$

the definition of the **autocorrelation function** of the observable $\widehat{f}$ is

$$\widehat{C}(t) = \widehat{C}_{ij} = \langle \, (f_i - \langle f_i \rangle) \, (f_j - \langle f_j \rangle) \, \rangle = \langle f_i f_j \rangle - \langle f_i \rangle \, \langle f_j \rangle = \langle f_0 f_t \rangle - \widehat{f}^{\,2} \tag{34}$$

where we used that translation invariance in time holds for the equilibrium ensemble. The asymptotic behavior for large t is

$$\widehat{C}(t) \sim \exp\left(-\frac{t}{\tau_{\exp}}\right) \quad \text{for} \quad t \to \infty, \tag{35}$$

where $\tau_{\exp}$ is called **exponential autocorrelation time** and is related to the second largest eigenvalue λ_1 of the transition matrix by $\tau_{\exp} = -\ln \lambda_1$ under the assumption that f has a non-zero projection on the corresponding eigenstate. Superselection rules are possible so that different autocorrelation times reign for different operators.

The variance of f is a special case of the autocorrelations (34)

$$\widehat{C}(0) = \sigma^2(f) \ . \tag{36}$$

Some algebra [3] shows that the variance of the estimator $\overline{f}$ (33) for the mean and the autocorrelation function (34) are related by

$$\sigma^2(\overline{f}) \;=\; \frac{\sigma^2(f)}{N}\left[1+2\sum_{t=1}^{N-1}\left(1-\frac{t}{N}\right)\widehat{c}(t)\right] \quad \text{with} \quad \widehat{c}(t)=\frac{\widehat{C}(t)}{\widehat{C}(0)}\;. \tag{37}$$

This equation ought to be compared with the corresponding equation for uncorrelated random variables: $\sigma^2(\overline{f}) = \sigma^2(f)/N$. The difference is the factor in the bracket of (37), which defines the **integrated autocorrelation time**

$$\tau_{\rm int} \;=\; \left[1+2\sum_{t=1}^{N-1}\left(1-\frac{t}{N}\right)\widehat{c}(t)\right]\;. \tag{38}$$

For correlated data the variance of the mean is by the factor $\tau_{\rm int}$ larger than the corresponding variance for uncorrelated data. In most simulations one is interested in the limit $N \to \infty$ and equation (38) becomes

$$\tau_{\rm int} \;=\; 1+2\sum_{t=1}^{\infty}\widehat{c}(t)\;. \tag{39}$$

The numerical estimation of the integrated autocorrelation time faces difficulties. The variance of the estimator for (39) diverges, because for large t each $\overline{c}(t)$ adds a constant amount of noise, whereas the signal dies out like $\exp(-t/\tau_{\rm exp})$. To obtain an estimate one considers the t-dependent estimator

$$\overline{\tau}_{\rm int}(t) \;=\; 1+2\sum_{t'=1}^{t}\overline{c}(t') \tag{40}$$

and looks out for a **window** in t for which $\overline{\tau}_{\rm int}(t)$ is flat.

3.2 Integrated autocorrelation time and binning

Using binning (also called blocking) the integrated autocorrelation time can also be estimated via the variance ratio. We bin the time series (32) into $N_{bs} \le N$ bins of

$$N_b = \texttt{NBIN} = \left[\frac{N}{N_{bs}}\right] = \left[\frac{\texttt{NDAT}}{\texttt{NBINS}}\right] \tag{41}$$

data each. Here [.] stands for Fortran integer division, *i.e.*, $N_b = \texttt{NBIN}$ is the largest integer $\le N/N_{bs}$, implying $N_{bs}\cdot N_b \le N$. It is convenient to choose the values of N and N_{bs} so that N is a multiple of N_{bs}. The binned data are the averages

$$f_j^{N_b} = \frac{1}{N_b}\sum_{i=1+(j-1)N_b}^{jN_b} f_i \quad \text{for} \;\; j=1,\ldots,N_{bs}\;. \tag{42}$$

For $N_b > \tau_{\rm exp}$ the autocorrelations are essentially reduced to those between nearest neighbor bins and even these approach zero under further increase of the binsize.

For a set of N_{bs} binned data $f_j^{N_b}$, $(j = 1, \ldots, N_{bs})$ we may calculate the mean with its naive error bar. Assuming for the moment an infinite time series, we find the integrated autocorrelation time (38) from the following ratio of sample variances

$$\tau_{\rm int} \;=\; \lim_{N_b \to \infty} \tau_{\rm int}^{N_b} \quad \text{with} \quad \tau_{\rm int}^{N_b} \;=\; \left(\frac{s^2_{\overline{f}^{N_b}}}{s^2_{\overline{f}}} \right) \; . \tag{43}$$

In practice the $N_b \to \infty$ limit will be reached for a sufficiently large, finite value of N_b. The statistical error of the $\tau_{\rm int}$ estimate (43) is, in the first approximation, determined by the errors of $s^2_{\overline{f}^{N_b}}$. The typical situation is then that, due to the central limit theorem, the binned data are approximately Gaussian, so that the **error of $s^2_{\overline{f}^{N_b}}$ is analytically known** from the χ^2 distribution. Finally, the fluctuations of $s^2_{\overline{f}}$ of the denominator give rise to a small correction which can be worked out [3].

Numerically most accurate estimates of $\tau_{\rm int}$ are obtained for the finite binsize N_b which is just large enough that the binned data (42) are practically uncorrelated. While the Student distribution shows that the confidence intervals of the error bars from 16 uncorrelated normal data are reasonable approximations to those of the Gaussian standard deviation, about 1000 independent data are needed to provide a decent estimate of the corresponding variance (at the 95% confidence level with an accuracy of slightly better than 10%). It makes sense to work with error bars from 16 binned data, but the error of the error bar, and hence a reliable estimate of $\tau_{\rm int}$, requires far more data.

3.3 Comparison of MCMC algorithms

Figure 2 illustrates $2d$ Ising model simulations on the critical point of its second order phase transition, $\beta = \beta_c = \ln(1 + \sqrt{2})/2$. **Critical slowing down** is observed: An increase $\tau_{\rm int} \sim L^z$ with lattice size, where $z \approx 2.17$ is the **dynamical critical exponent** of the $2d$ Ising model. Estimates of z are compiled in [22]. Using another MC dynamics the critical slowing down can be overcome by cluster updating [23, 24].

Figure 3 exhibits the **improvements of heat bath over Metropolis updating** for the 10-state $d = 2$ Potts model at $\beta = 0.62$. For this first order phase transition there is practically no lattice size dependence of the integrated autocorrelation time, once the lattices are large enough. We see that the 2-hist Metropolis updating reduces $\tau_{\rm int}$ by about a factor of two and the heatbath updating reduces it by about a factor of five.

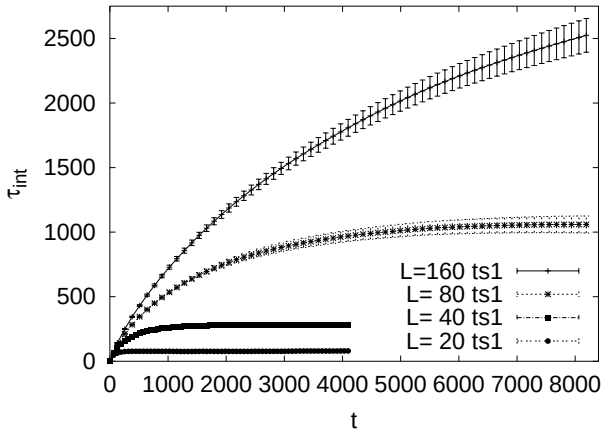

Fig. 2. One-hit Metropolis algorithm with sequential updating: Lattice size dependence of the integrated autocorrelation time for the $d = 2$ Ising model at its critical temperature (assignment `a0402_02` D of [3]). The ordering of the curves is identical with the ordering of the labels in the figure.

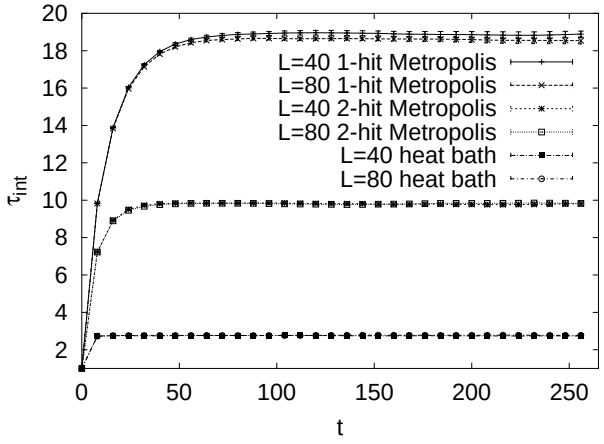

Fig. 3. Systematic updating: Comparison of the integrated autocorrelation times of the 1-hit and 2-hit Metropolis algorithms and the heat bath algorithm for the 10-state Potts model on $L \times L$ lattices at $\beta = 0.62$ (assignment `a0402_06`). The $L = 40$ and $L = 80$ curves lie almost on top of one another.

3.4 Jackknife estimators

Often one wants to estimate a non-linear function of the mean $\widehat{x}$ of some statistical variables $\widehat{f} = f(\widehat{x})$ where the estimator of $\widehat{x}$ and $\widehat{f}$ are

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \,, \qquad \overline{f} = f(\overline{x}) \;. \tag{44}$$

Typically, the bias is of order $1/N$:

$$\text{bias } (\overline{f}) \;=\; \widehat{f} - \langle \overline{f} \rangle \;=\; \frac{a_1}{N} + \frac{a_2}{N^2} + O(\frac{1}{N^3}) \tag{45}$$

where a_1 and a_2 are constants. But for the biased estimator we lost the ability to estimate the variance $\sigma^2(\overline{f}) = \sigma^2(f)/N$ via the standard equation

$$s^2(\overline{f}) \;=\; \frac{1}{N} s^2(f) \;=\; \frac{1}{N\,(N-1)} \sum_{i=1}^{N} (f_i - \overline{f})^2 \;, \tag{46}$$

because $f_i = f(x_i)$ is not a valid estimator of $\widehat{f}$. The error bar problem for the estimator $\overline{f}$ is conveniently overcome by using **jackknife estimators** $\overline{f}^J$, f_i^J, defined by

$$\overline{f}^J \;=\; \frac{1}{N}\sum_{i=1}^{N} f_i^J \;\text{ with }\; f_i^J \;=\; f(x_i^J) \;\text{ and }\; x_i^J \;=\; \frac{1}{N-1}\sum_{k\neq i} x_k \;. \tag{47}$$

The estimator for the variance $\sigma^2(\overline{f}^J)$ is

$$s_J^2(\overline{f}^J) \;=\; \frac{N-1}{N} \sum_{i=1}^{N} (f_i^J - \overline{f}^J)^2 \;. \tag{48}$$

Straightforward algebra shows that in the unbiased case the estimator of the jackknife variance (48) reduces to the normal variance (46). Notably, only order N (not N^2) operations are needed to construct the jackknife averages $x_i^J,\ i = 1, \ldots, N$ from the orginal data.

The jackknife method was introduced in the 1950s [25, 26]. For a review see [3]). It is recommended as the standard for error bar calculations of biased estimators.

3.5 Self-consistent versus reasonable error analysis

By visual inspection of the time series, one may get an impression about the length of the out-of-equilibrium part of the simulation. On top of this one should still choose

$$\texttt{nequi} \gg \tau_{\rm int} \ , \tag{49}$$

to allow the system to settle. That is a first reason, why it appears necessary to control the integrated autocorrelation time of a MC simulations. A second reason is that we have to control the error bars of the equilibrium part of our simulation. Ideally the error bars are calculated as

$$\triangle\overline{f} \ = \ \sqrt{\sigma^2(\overline{f})} \quad \text{with} \quad \sigma^2(\overline{f}) \ = \ \tau_{\rm int}\,\frac{\sigma^2(f)}{N} \ . \tag{50}$$

This constitutes a **self-consistent error analysis** of a MC simulation.

However, the calculation of the integrated autocorrelation time may be out of reach. Many more than the about twenty independent data are needed, which according to the Student distribution are sufficient to estimate mean values with reasonably reliable error bars.

In practice, one has to be content with what can be done. **Often this means to rely on the binning method.** We simply calculate error bars of our ever increasing statistics with respect to a fixed number of

$$\texttt{NBINS} \ \geq \ 16 \ . \tag{51}$$

In addition, we may put 10% of the initially planned simulation time away for reaching equilibrium. *A-posteriori*, this can always be increased. Once the statistics is large enough, our small number of binned data become effectively independent and our error analysis is justified.

How do we know that the statistics has become large enough? In practical applications there can be indirect arguments, like FSS estimates, which tell us that the integrated autocorrelation time is in fact (much) smaller than the achieved bin length. This is no longer self-consistent, as we perform no explicit measurement of $\tau_{\rm int}$, but it is a **reasonable error analysis**.

4 Generalized Ensembles for MCMC Simulations

The MCMC schemes discussed so far simulate the Gibbs canonical ensemble. Mean values of physical observables at the temperature chosen are obtained as arithmetic averages of the measurements. However, in statistical physics one would like to control the partition function, which allows to calculate observables at all temperatures and for the the proper normalization of the entropy and free energy. Also configurations, which are rare in the canonical, but well represented in another ensemble can be of physical interest. Finally the efficiency of the Markov process, i.e., the computer time needed for the convergence of an estimator of a physical quantity to a desired accuracy can depend greatly on the ensemble in which the simulations are performed.

4.1 Reweighting of the canonical ensemble

A first attempt to calculate the partition function by MCMC simulations dates back to a 1959 paper by Salsburg et al. [27]. As was already noticed by the authors their method is restricted to very small lattices. The reason is that their approach relies on what one calls in the modern language **reweighting** of the Gibbs canonical ensemble. It extrapolates data from a canonical simulation at one temperature to the canonical ensemble at another temperature.

The reweighting method has a long history. McDonald and Singer [28] were the first to use the equations of [27] to evaluate physical quantities over a range of temperatures. But thereafter the method was essentially forgotten and a recovery in lattice gauge theory [29, 30] focused on calculating complex zeros of the partition function. It remained to the paper by Ferrenberg and Swendsen [31], to formulate crystal-clear for what the method is particularly good, and for what not: It allows to focus on peaks of appropriate observables in the thermodynamic scaling limit, but it does not allow to cover a finite temperature range in the infinite volume limit. Off critical points, the reweighting range $\triangle\beta$ in $\beta = 1/(kT)$ decreases like $\triangle\beta \sim 1/\sqrt{N}$, where N is the number of degrees of freedom, which parametrizes the size of the system (e.g., the number of atoms). This follows from the fact that the energy is an extensive thermodynamic quantity, $E \sim N$ with fluctuations $\sim \sqrt{N}$. As β multiplies the energy, the change stays within the fluctuations as long as $\triangle\beta\, N \sim \sqrt{N}$, so that $\triangle\beta \sim 1/\sqrt{N}$ follows.

At second order phase transitions the reweighting range actually increases, because critical fluctuations are larger than non-critical fluctuations. One has then $\triangle E \sim N^x$ with $1/2 < x < 1$ and the requirement $\triangle\beta\, N \sim N^x$ yields $\triangle\beta \sim N^{x-1}$. For first order phase transitions one has a latent heat $\triangle E \sim N$, but this does not mean that the reweighting range becomes of order one. In essence, the fluctuations collapse, because the two phases become separated by an interface. One is back to fluctuations within either of the two phases where $\triangle\beta \sim 1/\sqrt{N}$ holds.

To estimate the partition function over a finite range $\triangle e$ in the energy density $e = E/N$, i.e., $\triangle E \sim N$, one can patch the histograms from canonical simulations at several temperatures. Such **multi-histogram methods** have also a long tradition too. In 1972 Valleau and Card [32] proposed the use of overlapping bridging distributions and called their method "multistage sampling". Free energy and entropy calculations become possible when one can link the temperature region of interest with a point in configuration space for which exact values of these quantities are known. Ref. [31] stimulated a renaissance of this approach. Modern work [33, 34, 35] developed efficient techniques to combine the overlapping distributions into one estimate of the spectral density and to control the statistical errors of the estimate. However, the patching of histograms from canonical simulations faces a number of limitations:

1. The number of canonical simulations diverges like $\sqrt{N}$ when one wants to cover a finite, non-critical range of the energy density.
2. At first order phase transition point, the canonical probability of configuration with an interface decreases $\sim \exp(-f_s\, A)$. Here f_s is the interfacial surface tension and A the minimal area of an interface, which divides the system into subsystems of distinct phases. For a system of volume L^d the area A diverges $\sim L^{d-1}$ in the infinite volume limit $L \to \infty$.

4.2 Generalized ensembles

One can cope with the difficulties of multi-histogram methods by allowing arbitrary sampling distributions instead of just the Gibbs-Boltzmann ensemble. This was first recognized by Torrie and Valleau [4] when they introduced **umbrella sampling**. However, for the next thirteen years application of this idea remained mainly confined to physical chemistry. That there is a potentially very broad range of applications for the basic idea remained unrecognized. A major barrier, which prevented researchers from trying such extensions, was certainly the apparent lack of direct and straightforward ways of determining suitable weighting functions for problems at hand. In the words of Li and Scheraga [36]: *The difficulty of finding such weighting factors has prevented wide applications of the umbrella sampling method to many physical systems.*

The turn-around came with the introduction of the **multicanonical ensemble** [7, 8, 9]. These papers focused on one well-defined weight function, up to a normalization constant the inverse spectral density,

$$w_{\rm muca}(E) \sim 1/n(E) \quad \text{for} \quad E_{\rm min} \le E \le E_{\rm max}\,, \tag{52}$$

where $n(E)$ is the number of states, and offered a variety of methods to find a **"working approximation"** of the weight function. Here a working approximation is defined as being accurate enough, so that the desired energy range will indeed be covered after the weight factors are fixed. A typical multicanonical simulation consists then of three parts:

1. Construct a working approximation of the weight function $w_{\rm muca}$.
2. Perform a conventional MCMC simulation with these weight factors.
3. Reweight the data to the Gibbs-Boltzmann ensemble at desired temperatures to obtain estimates of canonical expectation values for observables of physical of interest.

For instance, for the statistical physics system considered in [8], the $2d$ 10-state Potts model, finite size scaling consideration allow to construct the weight function on a larger lattice in one step from the information about the spectral density on the already simulated smaller lattices. This is a solution to step (1) in this particular case. The simulation and data analysis (see [3] for details) is then rather straightforward. Still, such conceptual simplifications

might have changed little on the situation that non-canonical ensembles were rarely used, if there were not other favorable circumstance.

One was that the paper by [8] estimated the interfacial tension of the $2d$ 10-state Potts model and produced a result, which was an order of magnitude smaller than previous numerical estimates by renown computational scientists. Normally this would have been followed by an extended debate of the advantages and disadvantages of the two competing messages. However, shortly after publication of the numerical estimates it turned out that the interfacial tension of the $2d$ 10-state Potts models can be calculated analytically [18], and the multicanonical estimate was within 3% of the rigorous result. This attracted attention and gave a lot of researchers confidence in the method.

Another phenomenon was, that at 1991 ± 5 years a number of papers [5, 6, 11, 10, 12] emerged in the literature, which all aimed at improving MCMC calculations by extending the confines of the canonical ensemble. Most important has been the **replica exchange method**, which is also known under the names **parallel tempering** and **multiple Markov chains**. In the context of spin glass simulations an exchange of partial lattice configurations at different temperatures was proposed by Swendsen and Wang [5]. But it was only in the later works [6, 12], essentially by rediscovery, recognized that the special case for which entire configurations at distinct temperatures are exchanged is of utmost importance.

In the successful replica exchange method one performs n canonical MC simulations at different β-values with Boltzmann weight factors

$$w_{B,i}(E^{(k)}) = e^{-\beta_i E^{(k)}} = e^{-H}\,,\ i = 0, \ldots, n-1 \tag{53}$$

where $\beta_0 < \beta_1 < ... < \beta_{n-2} < \beta_{n-1}$, and exchanges neighboring β-values

$$\beta_{i-1} \longleftrightarrow \beta_i \quad \text{for} \ \ i = 1, \ldots, n-1\ . \tag{54}$$

Their joint Boltzmann factor is

$$e^{-\beta_{i-1}E_{i-1} - \beta_i E_i} = e^{-H} \tag{55}$$

and the $\beta_{i-1} \leftrightarrow \beta_i$ exchange leads to

$$\begin{aligned} -\triangle H &= \left(-\beta_{i-1}E_i^{(k)} - \beta_i E_{i-1}^{(k')}\right) - \left(-\beta_i E_i^{(k)} - \beta_{i-1}E_{i-1}^{(k')}\right) \\ &= \ (\beta_i - \beta_{i-1})\ \left(E_i^{(k)} - E_{i-1}^{(k')}\right) \end{aligned} \tag{56}$$

which is accepted or rejected according to the Metropolis algorithm, *i.e.*, with probability one for $\triangle H \leq 0$ and with probability $\exp(-\triangle H)$ for $\triangle H > 0$. The β_i-values should to be determined so that an appropriate large acceptance rate is obtained for **each** β exchange. This can be done by recursions [3], which are straightforward modifications of a method introduced in Ref. [37].

Finally, and perhaps most important: From about 1992 on applications of generalized ensemble methods diversified tremendously. This is documented in a number of reviews [38, 39, 40, 41]. In the next section we focus on the use of generalized ensembles in biophysics.

4.3 Generalized ensembles and biophysics

In Ref. [42] the multicanonical ensemble was first used for simulations of complex systems with frustrated interactions, in that case the Edwards-Anderson Ising spin glass. Multicanonical simulations of biologically relevant molecules followed [13, 14], in Ref. [14] under the name "entropic sampling", but this is just identical with multicanonical sampling [43].

The interactions defined by an energy function are frustrated if one cannot simultaneously align all variables favorably with respect to their mutual interactions. So one gets "frustrated", a situation well known to political decision makers. In physical models frustrated interactions lead to a rugged free energy landscape. Canonical MCMC simulations tend to get stuck in the free energy barriers. In a nutshell, another ensemble may smoothen out such barriers, jump them, or at least allow to escape from them, for instance through a disordered phase. This can accelerate the over-all convergence of the MCMC process considerably. If all parameters are chosen right, reweighting will finally allow to reconstruct canonical ensemble expectation values at desired temperatures.

The parallel tempering (replica exchange) method was introduced in Ref. [44] to the simulation of biomolecules. In particular its extension to Molecular Dynamics [45] has subsequently been tremendously successful. Nowadays folding of small proteins is achieved using PC clusters and it appears that all these simulations rely on some form of the replica exchange method.

4.4 Overcoming free energy barriers

Basic mechanisms for overcoming energy barriers are best illustrated for first-order phase transitions. There one deals with the simplified situation of a single barrier, which is easily visualized by plotting histograms of an appropriate observable. To give a first example, Fig. 4 shows for the $2d$ 10-state Potts model the canonical energy histogram at a pseudo-critical temperature versus the energy histogram of a multicanonical simulation on a 70×70 lattice [8]. The energy barrier is overcome by enhancing the probabilities for configurations, which are suppressed in the canonical ensemble due to an interfacial tension.

The same barrier can be overcome by a parallel tempering simulation, but in a quite different way. Fig. 5 shows the histograms from a parallel tempering simulation with eight processes on 20×20 lattices. Each histogram corresponds to a fixed temperature, given by the β values in the figure. The β_4 and β_5 values produce the clearest double peak histograms. For β_4 the

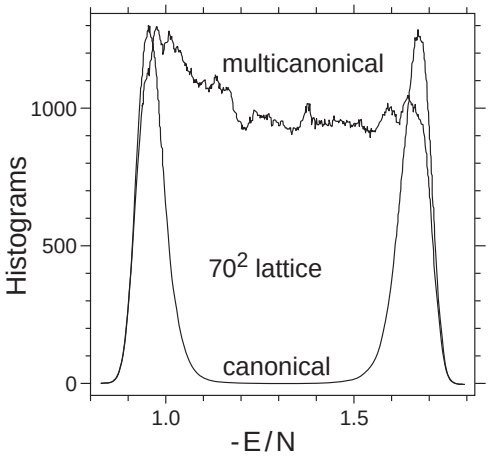

Fig. 4. Multicanonical $P_{mu}(E)$ together with canonical $P(E)$ energy distribution as obtained in Ref.[8] for the $2d$ 10-state Potts model on a 70×70 lattice.

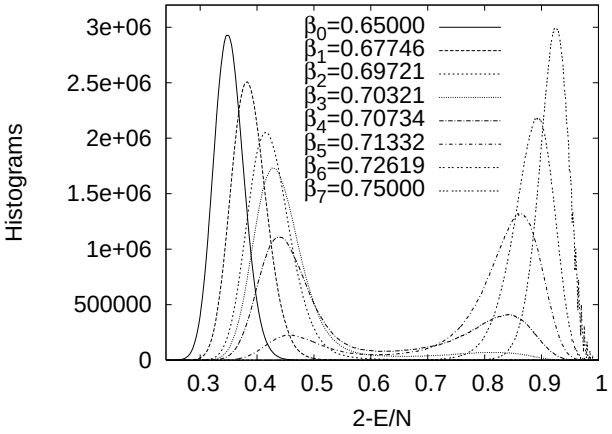

Fig. 5. Energy histograms from a parallel tempering simulation with eight processes for the $2d$ 10-state Potts model on 20×20 lattices (assignment `a0603_04` in [3]).

higher peak is in the disordered region and for β_5 it is in the ordered region. So the barrier can be "jumped" when there are at least two temperatures in the ensemble, which are sufficiently close to the particular pseudo-critical temperature for which the two peaks of the histogram assume equal heights (pseudocritical temperatures may also be defined by giving equal weights to the two parts of the histogram [46]). One of these two temperatures has to be in the ordered, the other in the disordered phase, and their start configurations have to be in the corresponding phases. The barrier can be jumped by an exchange of these two temperatures. If the barrier is high enough, so

that a single canonical simulations does not cross it during the simulation time, the jumping process determines the relative height of the two barriers. Adding more close-by temperatures to the ensemble will increase to the accuracy. Additional complications can occur, if a rare canonical tunneling process (crossing of the barrier) takes actually place.

Let us compare with a multicanonical simulation. The multicanonical method flattens the barrier, whereas the parallel tempering simulation allows to jump it. Using larger lattices the multicanonical method is well suited for calculations of the interface tensions from energy histograms [8]. For parallel tempering this is not the case, because the sampling of each histogram is still canonical. This can be overcome by a Gaussian variant of simulated tempering [47].

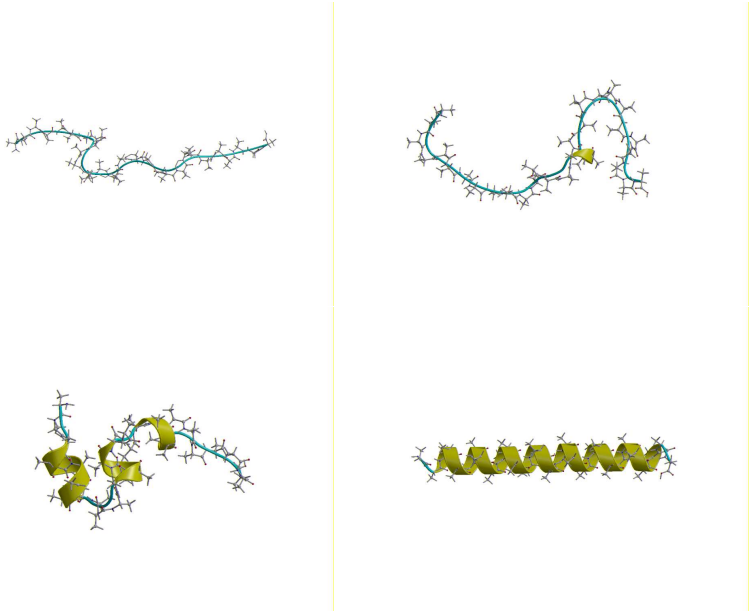

Fig. 6. Example configurations from a multicanonical simulation of poly-alanine [48] (courtesy Ulrich Hansmann and Yuko Okamoto).

In complex systems with a rugged free energy landscape the barriers can no longer be explicitly controlled. Nevertheless it has turned out that switching to the discussed ensembles can greatly enhance the MCMC efficiency. A variety of biologically relevant applications are reviewed in Ref. [39, 41] and in some of the lectures of this volume. Here we confine ourselves to showing a particularly nice example in Fig. 6: The folding of poly-alanine into its α-helix coil [48]. No *a-priori* information about the groundstate conformation is used in these kind of simulations. Whereas in a canonical simulation one would in essence

get stuck in a configuration like the second of the first row of this figure, the multicanonical simulation finds its way in and out of the helix structure.

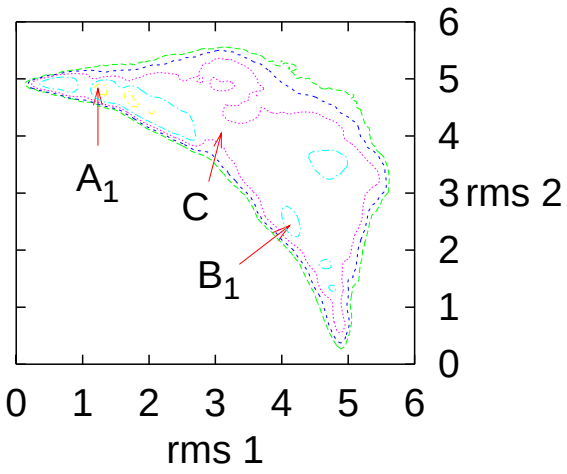

Fig. 7. Free energy landscape of Met-Enkephalin at $T = 250$ K with respect to rms distances (Å) from two reference configurations [51].

Variations of the basic idea introduce weights in other variables than the energy (or even in several variables). For instance, weights in the magnetic field of spin systems were introduced quite some while ago [49]. For spin glasses weights in the Parisi overlap variable are used [50]. The overlap of the configuration at hand with two reference configurations allows one to determine the transition state efficiently [51]. Fig. 7 shows a transition state located by this method in the free energy landscape of a simple peptide (Met-Enkephalin, which is discussed in the next section). In this figure contour lines are drawn every $2k_BT$. The labels A_1 and B_1 indicate the positions for the local-minimum states that originate from the reference configuration 1 and the reference configuration 2, respectively. The label C stands for the saddle point that corresponds to the transition state.

4.5 Weight factor recursions

For systems with rugged free energy landscapes FSS methods for estimating the weight factors are normally not available, because the systems are either of one particular size (e.g., biomolecules) or change too much when the system size is increased (e.g., spin glasses). In view of this one has to rely on ad-hoc per hand estimates or, more conveniently, on general purpose recursions. Designs for the latter were reported in a number of papers [52, 53, 54, 55].

The recursions have in common that they intend to create weight factors which are inversely proportional to the spectral density of the system at hand (modifications to target other weights are straightforward). The Wang-Landau (WL) recursion [55, 56] differs fundamentally from the other approaches by iterating the weight factor at energy E multiplicatively instead of additively. At a first glance the WL approach is counter-intuitive, because the correct iteration of the weight factor after a sufficiently long simulation is obviously proportional to one over the number of histogram entries $H(E)$ and not $1/(f_{WL})^{H(E)}$ with $f_{WL} > 1$. The advantage of the WL recursion is that it moves right away rapidly through the targeted energy range. Once this energy range is sufficiently covered, the WL method refines its iteration factor by $f_{WL} \to \sqrt{f_{WL}}$, so that it approaches 1. In this way the working approximations to the desired weight factors can be obtained on any finite energy range. In the following we give details for a variant [3] of the multicanonical recursion [52] (the modifications rely on work with W. Janke) and the WL recursion [55].

The multicanonical recursion uses the parameterization [7, 42]

$$w(a) \;=\; e^{-S(E_a)} \;=\; e^{-b(E_a)\,E_a + a(E_a)}$$

of the weights, where (for ϵ smallest stepsize)

$$\begin{aligned} b(E) &= \left[S(E+\epsilon) - S(E)\right]/\epsilon \quad \text{and} \\ a(E-\epsilon) &= a(E) + \left[b(E-\epsilon) - b(E)\right]\, E \;. \end{aligned}$$

After some algebra and other consideration [3] the recursion reads

$$\begin{aligned} b^{n+1}(E) &= b^n(E) + \hat{g}_0^n(E)\,[\ln H^n(E+\epsilon) - \ln H^n(E)]/\epsilon\,, \\ \hat{g}_0^n(E) &= g_0^n(E)\,/\,[g^n(E) + \hat{g}_0^n(E)]\,, \\ g_0^n(E) &= H^n(E+\epsilon)\,H^n(E)\,/\,[H^n(E+\epsilon) + H^n(E)]\,, \\ g^{n+1}(E) &= g^n(E) + g_0^n(E),\;\; g^0(E) = 0\,. \end{aligned}$$

For continuous systems like biomolecules some modification are required, see [57].

For the WL recursion updates are performed with estimators $\rho(E)$ of the density of states

$$p(E_1 \to E_2) = \min\left[\frac{\rho(E_1)}{\rho(E_2)}, 1\right] \;.$$

Each time an energy level is visited, they update the estimator

$$\rho(E) \to \rho(E)\, f_{WL}$$

where, initially, $\rho(E) = 1$ and $f_{WL} = f_0 = e^1$. Once the desired energy range is covered, the factor f_{WL} is refined to

$$f_1 = \sqrt{f},\ f_{n+1} = \sqrt{f_{n+1}}$$

until some small value like $f_{WL} = e^{-8} = 1.00000001$ is reached. For f_{WL} very close to one the difference to a simulation with fixed weights becomes negligible, so that one may just keep on iterating $f_{WL} \to \sqrt{f_{WL}}$. However, it appears that such a continued iteration is far less efficient than switching to a simulation with fixed weights as soon as a working estimate is found. Surprisingly there appears to be only one comparative study [58] of the different recursions, which finds that overall performances are similar.

5 Biased Markov Chain Monte Carlo

Consider a random variable y which is sampled with a probability density function (PDF) $P(y)$ on an interval $[y_1, y_2]$. The cumulative distribution function (CDF) is defined by

$$z = F(y) = \int_{y_1}^{y} P(y')dy' \quad \text{and} \quad P(y) = \frac{dF(y)}{dy}\,, \tag{57}$$

where $P(y)$ is properly normalized so that $F(\infty) = 1$ holds.

The HBA generates y by converting a uniformly distributed random number $0 \le z < 1$ into

$$y = F^{-1}(z)\,. \tag{58}$$

As the *acceptance rate* is defined by the number of accepted changes divided by the total number of proposed moves, the acceptance rate of the HBA is always 1 (a new value of y is generated on *every* step). In simulations the inversion of the CDF (57) may be unacceptably slow or the CDF itself may not be *a priori* known. Then one has to rely on other approaches.

In the conventional Metropolis scheme y_{new} is generated uniformly in a range $[y_1, y_2]$ (we refer to this as **proposal**) and accepted with probability (**accept/reject step**)

$$p_{Met} = \min\left\{1, \frac{P(y_{new})}{P(y_{old})}\right\}. \tag{59}$$

This process can have a low acceptance rate in the region of interest. Possible remedies are to decrease the proposal range, which makes the moves small, or use of multi-hit Metropolis. Excluding CPU time considerations for the

moment, measured by the integrated autocorrelation time both remedies are normally less efficient than a HBA.

Hastings [59] identified proposal probabilities, which are more general than those of the conventional Metropolis scheme, but gave no guidance whether some probabilities may be preferable over others. If one does not propose y_{new} uniformly anymore, the name **Biased Metropolis Algorithm (BMA)** is often used. Some biased Metropolis simulations can be found in the literature where the bias is introduced to match special situations [60, 61, 62, 63, 64]. In the following we discuss a general biased Metropolis scheme [15, 65], which aims at approximating heatbath probabilities.

Let us discretize y into n bins as

$$y_1 = y^0 < y^1 < y^2 < ... < y^n = y_2 \tag{60}$$

where the lengths of the bins are

$$\triangle y^j = y^j - y^{j-1}, \quad \text{with} \quad j = 1, ..., n. \tag{61}$$

A BMA can then be defined by the following steps:

- Propose a new value y_{new} by randomly picking a bin j_{new} and then proposing y_{new} uniformly in the given bin. Two uniform random numbers r_1, r_2 are needed:

$$j_{new} = 1 + \mathrm{Int}[n\, r_1] \quad \text{and} \quad y_{new} = y^{j_{new}-1} + \triangle y^{j_{new}}\, r_2, \tag{62}$$

  where $\mathrm{Int}[n\, r_1]$ denotes rounding to the largest integer $\leq n\, r_1$.
- Locate the bin j_{old} to which y_{old} belongs:

$$y^{j_{old}-1} \leq y_{old} \leq y^{j_{old}}. \tag{63}$$

- Accept y_{new} with probability:

$$p_{BMA} = \min\left\{1, \frac{P(y_{new})}{P(y_{old})}\, \frac{\triangle y^{j_{new}}}{\triangle y^{j_{old}}}\right\} \tag{64}$$

p_{BMA} in (64) differs from p_{Met} in (59) by the *bias* $\triangle y^{j_{new}}/\triangle y^{j_{old}}$. The scheme outlined in (62)-(64) satisfies balance or detailed balance in the same way as the original Metropolis algorithm, while the bias changes the acceptance rate.

So far the partitioning y^j has not been introduced explicitly. A particular choice is:

$$\frac{j}{n} = F(y^j) \quad \text{or} \quad y^j = F^{-1}\left(\frac{j}{n}\right). \tag{65}$$

This equation achieves equidistant partitioning on the CDF ordinate. Let us pick a bin initially labeled j and take the limit $n \to \infty$ so that this bin collapses into a point labeled z. Then this BMA approaches the HBA and the acceptance rate converges to 1:

$$\frac{P(y_{new})}{P(y_{old})}\,\frac{\triangle y^{j_{new}}}{\triangle y^{j_{old}}} \rightarrow \frac{P(y_{new})}{P(y_{old})}\,\frac{1/P(y_{new})}{1/P(y_{old})} = 1. \tag{66}$$

Therefore we call a BMA with the partitioning (65) **Biased Metropolis-Heatbath Algorithm (BMA)**. Restricted to the updating of one dynamical variable the improvements are similar to those seen in Fig. 3, where the gain is a factor of five. Having in mind that that realistic simulations take usually months of computer time, such factors are highly welcome. Extending the biased method to simultaneous updating of more than one variable, larger improvement factors can be anticipated. But extensions to more than two variables face technical difficulties.

5.1 Application to a continuous model

Following Ref. [65] we illustrate the BMHA for a system with a continuous energy function: Pure lattice gauge theory calculations with the $U(1)$ gauge group for which the "matrices" are complex numbers on the unit circle, parameterized by an angle $\phi \in [0, 2\pi)$. After defining the theory on the links of a four-dimensional lattice and going through some algebra, the PDF

$$P_\alpha(\phi) = N_\alpha\, e^{\alpha\,\cos(\phi)} \tag{67}$$

has to be sampled, where α is a parameter associated to the interaction of the link being updated with its environment. The corresponding CDF is

$$F_\alpha(\phi) = N_\alpha \int_0^\phi d\phi'\, e^{\alpha\,\cos(\phi')} \tag{68}$$

where N_α ensures the normalization $F_\alpha(2\pi) = 1$. In the following we consider $U(1)$ gauge theory at a coupling close to the critical point for which one finds $0 \leq \alpha \leq 6$.

Let us discretize the parameter α into $m = 2^{n_1} = 16$ ($n_1 = 4$) bins, choosing equidistant partitioning. In each α^i bin we discretize ϕ using the condition (65) with $n = 2^{n_2} = 16$ ($n_2 = 4$). Two two-dimensional arrays are needed: one for storing $\phi^{i,j}$ (levels themselves) and another for $\triangle\phi^{i,j} = \phi^{i,j} - \phi^{i,j-1}$ (distances between levels), see Fig. 8. For a given α^i it is straightforward to apply BMA step (62). E.g., for $\alpha = \alpha^{11}$, the cross section of the $F_\alpha(\phi)$ surface plane is then shown in Fig. 9. To determine the bin label j_{old} which belongs to the (known) value $\phi_{0,old}$ (BMA step (63)) one may use the n_2-step recursion $j \;\rightarrow\; j + 2^{i_2}\ \mathrm{sign}\,(\,\phi - \phi^{i,j}\,)$, $i_2 \rightarrow i_2 - 1$. Once j_{old} is known it gives the length of the bin: $\triangle\phi^{i,j_{old}}$ and the final accept/reject step (64) can be applied:

$$p_{BMA} = \min\left\{1,\ \frac{\exp\left(\alpha\,\cos\phi_{0,new}\right)}{\exp\left(\alpha\,\cos\phi_{0,old}\right)}\,\frac{\triangle\phi_0^{i,j_{new}}}{\triangle\phi_0^{i,j_{old}}}\right\}. \tag{69}$$

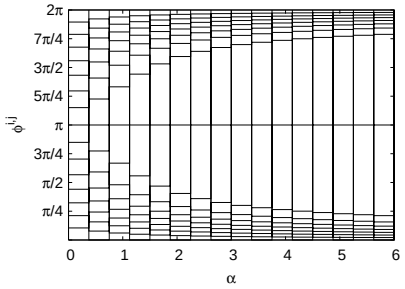

Fig. 8. $m \times n$ partitioning of $\triangle\phi^{i,j}$ for $U(1)$ at the coupling constant value discussed in the text.

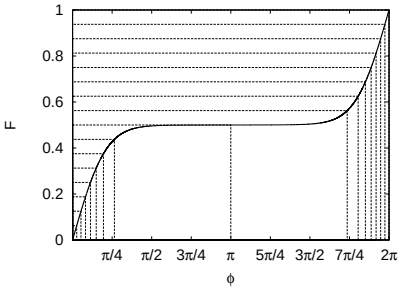

Fig. 9. Discretization of the CDF $F_{\alpha^{11}}(\phi)$ for $U(1)$ corresponding to the 11th bin of Fig. 8.

In this example the CDF is known. We have shown that sampling with the BMHA is essentially equivalent to using the HBA, but can be numerically faster, as is the case for $U(1)$. $SU(2)$ lattice gauge theory with the fundamental-adjoint action allows for substantial speed-ups by using a BMHA [66]. In the next section we show how a similar biasing procedure can be used when the CDF is not known (making a HBA impossible) and extend it to two variables.

5.2 Rugged Metropolis, a biasing scheme for biophysics

We consider biomolecule models for which the energy E is a function of a number of dynamical variables v_i, $i = 1, \ldots, n$. The fluctuations in the Gibbs canonical ensemble are described by a probability density function $\rho(v_1, \ldots, v_n; T) = const\ \exp[-\beta\, E(v_1, \ldots, v_n)]$, where T is the temperature, $\beta = 1/(kT)$, and E is the energy of the system. To be consistent with the notation of [15] we now use $\rho(v_1, \ldots, v_n; T)$ instead of $P(y)$ introduced in previous one-variable example. Proposing a new variable (with the other variables fixed) from the PDF constitutes a HBA. However, an implementation of a HBA is only possible when the CDF of the PDF can be controlled. In particular this requires the normalization constant in front of the $\exp[-\beta\, E(v_1, \ldots, v_n)]$ Boltzmann factor, which is normally unknown. Then the following strategy can provide a useful approximation.

For a range of temperatures

$$T_1 \;>\; T_2 \;>\; \ldots \;>\; T_r \;>\; \ldots \;>\; T_{f-1} \;>\; T_f \tag{70}$$

the simulation at the highest temperature, T_1, is performed with the usual Metropolis algorithm and the results are used to construct an *estimator*

$$\overline{\rho}(v_1, \ldots, v_n; T_1)$$

which is used to bias the simulation at T_2. Recursively, the estimated PDF

$$\overline{\rho}(v_1, \ldots, v_n; T_{r-1})$$

is expected to be a useful approximation of $\rho(v_1, \ldots, v_n; T_r)$. Formally this means that BMA acceptance step (64) at temperature T_r is of the form

$$P_{RM} = \min \left\{ 1, \frac{\exp(-\beta E')}{\exp(-\beta E)} \frac{\overline{\rho}(v_1, \ldots, v_n; T_{r-1})}{\overline{\rho}(v'_1, \ldots, v'_n; T_{r-1})} \right\} \tag{71}$$

where $\beta = 1/(kT)$. For this type of BMA, where the bias is constructed by using information from a higher temperature, the name **Rugged Metropolis (RM)** was given in [15].

Our test case in the following will is the small brain peptide Met-Enkephalin (Tyr-Gly-Gly-Phe-Met) in vacuum, which features 24 dihedral angels as dynamical variables. Its global energy minimum was determined some time ago by Li and Scheraga [67]. Ever since this molecule is often used as a simple laboratory for testing new computational methods. We rely on the all-atom energy function ECEPP/2 (Empirical Conformational Energy Program for Peptides) [68] as implemented in the SMMP (Simple Molecular Mechanics for Proteins) [69] program package. Besides the ϕ, ψ angles, we keep also the ω angles unconstrained, which are usually restricted to $[\pi - \pi/9, \pi + \pi/9]$. This allows us to illustrate the RM idea for a particularly simple case.

To get things started, we need to construct an estimator $\overline{\rho}(v_1, \ldots, v_n; T_r)$ from the numerical data of the RM simulation at temperature T_r. This is neither simple nor straightforward, and approximations have to be used.

The $\mathbf{RM_1}$ approximation

In Ref. [15] the factorization

$$\overline{\rho}(v_1, \ldots, v_n; T_r) = \prod_{i=1}^{n} \overline{\rho}_i^1(v_i; T_r) \tag{72}$$

was investigated, where $\overline{\rho}_i^1(v_i; T_r)$ are estimators of reduced one-variable PDFs defined by

$$\rho_i^1(v_i; T) = \int_{-\pi}^{+\pi} \prod_{j \neq i} d\, v_j\, \rho(v_1, \ldots, v_n; T) \ . \tag{73}$$

The resulting algorithm, called RM_1, constitutes the simplest RM scheme possible. The CDFs are defined by

$$F_i(v) \ = \ \int_{-\pi}^{v} dv'\, \rho_i^1(v') \ . \tag{74}$$

The estimate of F_{10}, the CDF for the dihedral angle Gly-3 ϕ (v_{10}), from the simulations at our highest temperature, $T_1 = 400\,\mathrm{K}$, is shown in Fig. 10. For

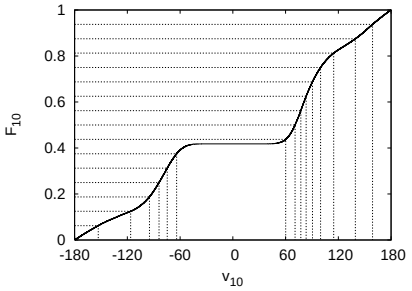

Fig. 10. Estimate of the cumulative distribution function for the Met-Enkephalin dihedral angle v_{10} (Gly-3 ϕ) at $400\,K$.

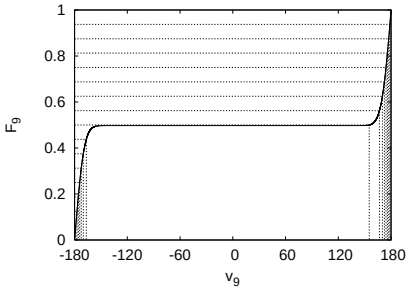

Fig. 11. Estimate of the cumulative distribution function for the Met-Enkephalin dihedral angle v_9 (Gly-2 ω) at 400 K.

our plots we use degrees, while we use radiant in our theoretical discussions and in the computer programs. Fig. 10 is obtained by sorting all $n_{\rm dat}$ values of v_{10} in our time series in ascending order and increasing the values of F_{10} by $1/n_{\rm dat}$ whenever a measured value of v_{10} is encountered. Using a heapsort approach [3], the sorting is done in $n_{\rm dat}\,\log_2(n_{\rm dat})$ steps.

Figure 11 shows the CDF for v_9 (Gly-2 ω) at $400\,K$, which is the angle of lowest acceptance rate in the conventional Metropolis updating. This distribution function corresponds to a histogram narrowly peaked around $\pm\pi$, which is explained by the specific electronic hybridization of the CO-N peptide bond. From the grid shown in Fig. 11 it is seen that the RM_1 updating concentrates the proposal for this angle in the range slightly above $-\pi$ and slightly below $+\pi$. Thus the procedure has automatically a similar effect as the often used restriction to the range $[\pi - \pi/9, \pi + \pi/9]$, which is also the default implementation in SMMP.

After the empirical CDFs are constructed for each angle v_i, they are discretized using the condition (65). Here we denote differences (61) needed for the bias by

$$\triangle v_{i,j} = v_{i,j} - v_{i,j-1} \quad \text{with} \quad v_{i,0} = -\pi \ . \tag{75}$$

The RM_1 updating of each dihedral angle v_i follows then the BMA procedure (62)-(64). The accept/reject step in the $v_{i,j}$ notation is

$$p_{RM_1} = \min\left\{1, \frac{\exp(-\beta E')}{\exp(-\beta E)}\,\frac{\triangle v_{i,j_{new}}}{\triangle v_{i,j_{old}}}\right\} \ . \tag{76}$$

The RM_2 approximation

In Ref. [70] the RM_1 scheme of Eq. (76) was generalized to the *simultaneous* updating of two dihedral angles. For $i_1 \neq i_2$ reduced two-variable PDFs are defined by

$$\rho^2_{i_1,i_2}(v_{i_1}, v_{i_2}; T) = \int_{-\pi}^{+\pi} \prod_{j \neq i_1,i_2} d\, v_j \, \rho(v_j, \ldots, v_n; T) \; . \tag{77}$$

The one-variable CDFs F_{i_1} and the discretization $v_{i_1,j}$, $j = 0, \ldots, n$ are already given by Eqs. (74) and (75). We define conditional CDFs by

$$F_{i_1,i_2;j}(v) = \int_{-\pi}^{v} dv_{i_2} \int_{v_{i_1,j-1}}^{v_{i_1,j}} dv_{i_1} \, \rho^2_{i_1,i_2}(v_{i_1}, v_{i_2}) \tag{78}$$

for which the normalization $F_{i_1,i_2;j}(\pi) = 1/n$ holds. To extend the RM_1 updating to two variables we define for each integer $k = 1, \ldots, n$ the value $F_{i_1,i_2;j,k} = k/n^2$. Next we define $v_{i_1,i_2;j,k}$ through $F_{i_1,i_2;j,k} = F_{i_1,i_2;j}(v_{i_1,i_2;j,k})$ and also the differences

$$\triangle v_{i_1,i_2;j,k} = v_{i_1,i_2;j,k} - v_{i_1,i_2;j,k-1} \quad \text{with} \quad v_{i_1,i_2;j,0} = -\pi \; . \tag{79}$$

The RM_2 procedure for the simultaneous update of (v_{i_1}, v_{i_2}) is then specified as follows:

- Propose a new value $v_{i_1,new}$ using two uniform random numbers r_1, r_2 (BMA step (62) for the angle i_1):

$$j_{new} = [n\, r_1] \quad \text{and} \quad v_{i_1,new} = v_{i_1,j_{new}-1} + \triangle v_{i_1,j_{new}}\, r_2. \tag{80}$$

- Propose a new value $v_{i_2,new}$ using two uniform random numbers r_3, r_4 (BMA step (62) for the angle i_2):

$$k_{new} = [n\, r_3] \quad \text{and} \quad v_{i_2,new} = v_{i_1,i_2;j_{new},k_{new}-1} + \triangle v_{i_1,i_2;j_{new},k_{new}}\, r_4. \tag{81}$$

- Find the bin index j_{old} for the present angle $v_{i_1,old}$ through $v_{i_1,j_{old}-1} \leq v_{i_1,old} \leq v_{i_1,j_{old}}$, just like for RM_1 updating (BMA step (63) for v_{i_1}).
- Find the bin index k_{old} for the present angle $v_{i_2,old}$ through $v_{i_1,i_2;j_{old},k_{old}-1} \leq v_{i_2,old} \leq v_{i_1,i_2;j_{old},k_{old}}$ (again step (63) but for v_{i_2}).
- Accept $(v_{i_1,new},\, v_{i_2,new})$ with the probability

$$p_{RM_2} = \min\left\{1, \frac{\exp(-\beta E')}{\exp(-\beta E)} \, \frac{\triangle v_{i_1,j_{new}}}{\triangle v_{i_1,j_{old}}} \, \frac{\triangle v_{i_1,i_2;j_{new},k_{new}}}{\triangle v_{i_1,i_2;j_{old},k_{old}}}\right\} \; . \tag{82}$$

As for RM_1, estimates of the conditional CDFs and the intervals $\triangle v_{i_1,i_2;j,k}$ are obtained from the conventional Metropolis simulation at 400 K. In the following we focus on the pairs (v_7, v_8), (v_{10}, v_{11}) and (v_{15}, v_{16}). These angles correspond to the largest integrated autocorrelation times of the RM_1 procedure and are expected to be strongly correlated with one another because they are pairs of dihedral angles around a C_α atom.

The bias of the acceptance probability given in Eq. (82) is governed by the areas

$$\triangle A_{i_1,i_2;j,k} = \triangle v_{i_1,j} \, \triangle v_{i_1,i_2;j,k} \; .$$

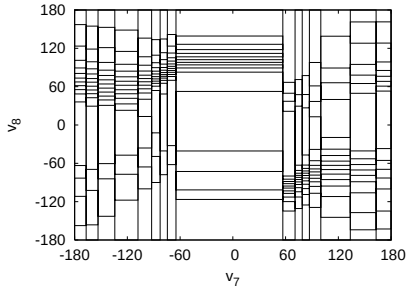

Fig. 12. Areas of equal probabilities (sorting v_7 then v_8).

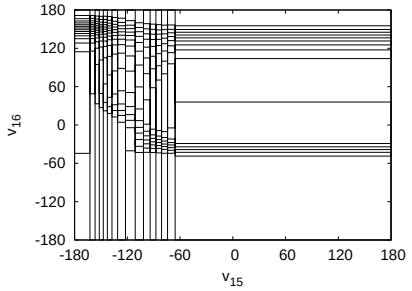

Fig. 13. Areas of equal probabilities (sorting v_{10} then v_{11}).

For $i_1 = 7$ and $i_2 = 8$ our 400 K estimates of these areas are depicted in Fig. 12. For the RM_2 procedure these areas take the role which the intervals on the abscissa of Fig. 10 play for RM_1 updating. The small and the large areas are proposed with equal probabilities, so the a-priori probability for our two angles is high in a small area and low in a large area. Areas of high probability correspond to allowed regions in the Ramachandran map of a Gly residue [71]. In Fig. 12 the largest area is 503.4 times the smallest area. Note that the order in which the angles are sorted introduces some differences. Fig. 13 gives a plot for (v_{15}, v_{16}) pairs in which the angle with the smaller subscript is sorted first. The ratio of the largest area over the smallest area is 2565.8. The large number is related to the fact that (v_{15}, v_{16}) is the pair of ϕ, ψ angles around the C_α atom of Phe-4, for which positive ϕ values are disallowed [71].

Reductions in the integrated autocorrelations times of the angles vary and are again similar to those ovserved in Fig. 3 when moving from the ordinary Metropolis to the HBA.

6 Outlook and Conclusions

Spin systems, lattice gauge theory and biophysics models are certainly far apart in their scientific objectives. Nevertheless quite similar computational techniques allow for efficient MCMC simulations in either field. Cross-fertilization may go in both directions. For instance, generalized ensemble techniques propagated from lattice gauge theory [7] over statistical physics [42] into biophysics [13], while it appears that biased Metropolis techniques [60, 61, 62, 63, 15] propagate in the opposite direction [65].

It is rather straightforward to combine the biased techniques of section 5 with generalized ensembles, but work in this direction has just begun. In Ref. [15] and [72] combinations with parallel tempering have been studied and the improvements were approximately multiplicative. An implementation into

the multicanonical ensemble has recently been worked out [73] and applied in a study of the deconfining phase transition of $U(1)$ lattice gauge theory. Extension of the rugged MC approach to MD are also possible [74]. As this leads into ongoing research, it is a good point to conclude these lecture notes at this point.

References

1. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller: J. Chem. Phys. **21**, 1087 (1953)
2. J. Gubernatis (editor): *The Monte Carlo Method in the Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm* (AIP Conference Proceedings, Volume 690, Melville, NY 2003)
3. B.A. Berg: *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore 2004)
4. G.M. Torrie and J.P. Valleau: J. Comp. Phys. **23**, 187 (1977)
5. R.H. Swendsen, and J.-S. Wang: Phys. Rev. Lett. **57**, 2607 (1986)
6. G.J. Geyer: Markov Chain Monte Carlo Maximum Likelihood. In *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, ed by E.M. Keramidas (Interface Foundation, Fairfax, Virginia 1991) pp 156–163
7. B.A. Berg and T. Neuhaus: Phys. Lett. B **267**, 249 (1991)
8. B.A. Berg, and T. Neuhaus: Phys. Rev. Lett. **68**, 9 (1992)
9. B.A. Berg: Int. J. Mod. Phys. **3**, 1083 (1992)
10. A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkanov, and P.N. Vorontsov-Velyaminov: J. Chem. Phys. **96**, 1776 (1992)
11. E. Marinari and G. Parisi: Europhys. Lett. **19**, 451 (1992)
12. K. Hukusima and K. Nemoto: J. Phys. Soc. Japan **65**, 1604 (1996)
13. U.H. Hansmann and Y. Okamoto: J. Comp. Chem. **14**, 1333 (1993)
14. M.-H. Hao and H.A. Scheraga: J. Phys. Chem. **98**, 4940 (1994)
15. B.A. Berg: Phys. Rev. Lett. **90**, 180601 (2003)
16. F.Y. Wu: Rev. Mod. Phys. **54**, 235 (1982)
17. R.J. Baxter: J. Phys. C **8**, L445 (1973)
18. C. Borgs and W. Janke: J. Phys. I France **2**, 2011 (1992)
19. J. Propp and D. Wilson: DIMACS Series in Discrete Mathematics and Theoretical Computer Science (AMS) **41**, 181 (1998)
20. W.S. Kendall: Notes on Perfect Simulation. In *Markov Chain Monte Carlo: Innovations and Applications (Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore)*, ed by W.S. Kendall, F. Liang, J.-S. Wang (World Scientific, Singapore 2005)
21. G. Marsaglia, A. Zaman and W.W. Tsang: Stat. Prob. **8**, 35 (1990)
22. D.P. Landau and K. Binder: *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge 2000)
23. R.H. Swendsen and J.-S. Wang: Phys. Rev. Lett. **58**, 86 (1987)
24. U. Wolff: Phys. Rev. Lett. **62**, 361 (1989)
25. M.H. Quenouille: Biometrika **43**, 353 (1956)
26. J.W. Tukey: Ann. Math. Stat. **29**, 614 (1958)
27. Z.W. Salsburg, J.D. Jacobson, W.S. Fickett, and W.W. Wood: J. Chem. Phys. **30**, 65 (1959)

28. I.R. McDonald and K. Singer: Discussions Faraday Soc. **43**, 40 (1967)
29. M. Falcioni, E. Marinari, L. Paciello, G. Parisi, and B. Taglienti: Phys. Lett. B **108**, 331 (1982)
30. E. Marinari: Nucl. Phys. B **235**, 123 (1984)
31. A. Ferrenberg and R. Swendsen: Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1658 (1989)
32. J.P. Valleau and D.N. Card: J. Chem. Phys. **57**, 5457 (1972)
33. A.M. Ferrenberg and R.H. Swendsen: Phys. Rev. Lett. **63**, 1195 (1989)
34. N.A. Alves, B.A. Berg, and R. Villanova: Phys. Rev. B **41**, 383 (1990)
35. N.A. Alves, B.A. Berg, and and S. Sanielevici: Nucl. Phys. B **376**, 218 (1992)
36. Z. Li and H.A. Scheraga: J. Mol. Struct. (Theochem) **179**, 333 (1988)
37. W. Kerler and P. Rehberg: Phys. Rev. E **50**, 4220 (1994)
38. W. Janke: Physica A **254**, 164 (1998)
39. U.H. Hansmann and Y. Okamoto: Ann. Rev. Comp. Phys. **6**, 129 (1999)
40. B.A. Berg: Fields Institute Communications **26**, 1 (2000) [cond-mat/990923]
41. A. Mitsutake, Y. Sugita and Y. Okamoto, Biopolymers (Peptide Science) **60**, 96 (2001)
42. B.A. Berg and T. Celik: Phys. Rev. Lett. **69**, 2292 (1992)
43. B.A. Berg, U.H. Hansmann and Y. Okamoto: J. Phys. Chem. **99**, 2236 (1995)
44. U.H. Hansmann: Chem. Phys. Letters **281**, 140 (1997)
45. Y. Sugita and Y. Okamoto: Chem. Phys. Lett. **314**, 141 (1999)
46. C. Borgs and S. Kappler: Phys. Lett. A **171**, 37 (1992)
47. T. Neuhaus and J.S. Hager: Phys. Rev. E **74**, 036702 (2006).
48. U.H. Hansmann and Y. Okamoto: J. Chem. Phys. **110**, 1267 (1999)
49. B.A. Berg, U.H. Hansmann, and T. Neuhaus: Phys. Rev. B **47**, 497 (1993)
50. B.A. Berg and W. Janke: Phys. Rev. Lett. **80**, 4771 (1998)
51. B.A. Berg, H. Noguchi and Y.Okamoto: Phys. Rev. E **68**, 036126 (2003).
52. B.A. Berg: J. Stat. Phys. **82**, 323 (1996)
53. B.A. Berg, A. Billoire, and W. Janke: Phys. Rev. B **61**, 12143 (2000)
54. Y. Sugita and Y. Okamoto: Chem. Phys. Lett. **329**, 261 (2000)
55. F. Wang and D.P. Landau: Phys. Rev. Lett **86**, 2050 (2001)
56. F. Wang and D.P. Landau: Phys. Rev. E **64**, 056101 (2001)
57. F. Yasar, T. Celik, B. Berg and H. Meirovitch: J. Comp. Chem. **21**, 1251 (2000)
58. Y. Okamoto: pp 248 in Ref. [2], [cond-mat/0308119]
59. W.K. Hastings: Biometrica **57**, 97 (1970)
60. A.D. Bruce: J. Phys. A **18**, L873 (1985)
61. A. Milchev, D.W. Heermann, and K. Binder: J. Stat. Phys. **44**, 749 (1986)
62. M.W. Deem and J.S. Bader: Mol. Phys. **87**, 1245 (1996)
63. G. Favrin, A. Irbäck, and F. Sjunnesson: J. Chem. Phys. **114**, 8154 (2001)
64. J.P. Ulmschneider and W.L. Jorgensen: J. Chem. Phys. **118**, 4261 (2003)
65. A. Bazavov and B.A. Berg: Phys. Rev. D **71** 114506 (2005)
66. A. Bazavov, B.A. Berg, and U.M. Heller: Phys. Rev. D **72**, 117501 (2005)
67. Z. Li and H.A. Scheraga: Proc. Nat. Acad. Sci. USA, **85**, 6611 (1987)
68. M.J. Sippl, G. Nemethy, and H.A. Scheraga: J. Phys. Chem. **88**, 6231 (1984)
69. F. Eisenmenger, U.H. Hansmann, S. Hayryan, and C.-K. Hu: Comp. Phys. Commun. **138**, 192 (2001)
70. B.A. Berg and H.-X. Zhou: Phys. Rev. E. **72**, 016712 (2005)
71. G.E. Schultz and R.H. Schirmer: *Principle of Protein Structure*, Springer, New York, 1979.
72. B.A. Berg and H.-P. Hsu: Phys. Rev. E **69**, 026703 (2004)
73. A. Bazavov and B.A. Berg: Work in preparation.
74. D. Min, W. Yang, and B.A. Berg: Work in preparation.